Does Computer-Aided Instruction Improve Children's Cognitive and Noncognitive Skills?

Hirotake Ito, Keiko Kasai, Hiromu Nishiuchi, and Makiko Nakamuro*

This paper examines the causal effects of computer-aided instruction (CAI) on children's cognitive and noncognitive skills. We ran a clustered randomized controlled trial at five elementary schools with more than 1,600 students near Phnom Penh, Cambodia. After 3 months of intervention, we find that the average treatment effects on cognitive skills are positive and statistically significant, while hours of study were unchanged both at home and in the classroom. This indicates that CAI is successful in improving students' learning productivity per hour. Furthermore, we find that CAI raises students' subjective expectation to attend college in the future.

Keywords: clustered randomized controlled trial, computer-assisted instruction, noncognitive skills
JEL codes: I21, I25, I30

I. Introduction

The World Bank recently made reference to a "learning crisis" (World Bank 2017), arguing that a large proportion of students in developing countries are failing to acquire even foundational skills at school, for example, basic math that is required when buying and selling in markets, handling household budgets, or transacting with banks or other financial institutions (Hanushek and Woessmann 2016).

*Hirotake Ito (corresponding author): Graduate School of Media and Governance, Keio University. E-mail: itouhrtk@keio.jp; Keiko Kasai: School of International Development, University of East Anglia. E-mail: keikokasai131@gmail.com; Hiromu Nishiuchi: Graduate School of International Management, Yokohama City University. E-mail: hiromunishiuchi@gmail.com; Makiko Nakamuro: Faculty of Policy Management, Keio University. E-mail: makikon@sfc.keio.ac.jp. This study was conducted as part of the project "Research on the Improvement in Resource Allocation and Productivity among the Healthcare and Education Service Industries" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). We thank Hanamaru Lab, especially Kei Kawashima, Kodai Tokumaru, and Daiki Watanabe for their support of the experiment in Cambodia. We also thank the managing editor, the anonymous referee, and participants in the Asian Development Bank (ADB)-International Energy Agency (IEA) roundtable, Makoto Yano, Masayuki Morikawa, Kyoji Fukao, and Tomohiko Inui for helpful comments and suggestions. We also gratefully acknowledge the financial support received from the MEXT/JSPS KAKENHI Grant Number: 18H05314. The usual ADB disclaimer applies.

Asian Development Review, vol. 38, no. 1, pp. 98–118
https://doi.org/10.1162/adev_a_00159

While many low-income countries have rapidly increased primary school enrollments in recent decades, they often face substantial obstacles in avoiding a learning crisis. First, increases in primary school enrollments have occurred along with increases in education inputs, such as teachers and other school resources. However, any decline in per capita inputs will likely reduce the quality of primary education. Second, hiring high-quality teachers is difficult in many developing countries because they are paid less than other comparably qualified professionals, particularly in urban areas. Third, any substantial gap between the abilities of low- and high-achieving students makes it difficult for teachers to set their level of instruction appropriately. Such situations produce a mismatch between a teacher's level of instruction and students' level of proficiency (Glewwe and Muralidharan 2016).

New technologies offer promising ways to mitigate such problems in developing countries. Although computer access in classrooms does not improve students' learning, as shown in Barrera-Osorio and Linden (2009), well-designed computer-assisted learning (CAL) allows students to access high-quality instructional materials even in the presence of severe teacher shortages and learn at their own pace and proficiency. However, the empirical evidence on the effect of computer-aided instruction (CAI) is mixed. In India, CAI was found to improve student performance substantially, especially for low-achieving students (Linden 2008), while the One Laptop per Child programs in Peru and Uruguay had no impact on student reading or math abilities (Cristia et al. 2017; De Melo, Machado, and Miranda 2014).

This study was designed to rigorously estimate the causal impact of CAI on students' cognitive and noncognitive skills, in collaboration with the Government of Cambodia, the Japan International Cooperation Agency, and Hanamaru Lab, a Japanese private company that developed a personalized computer-assisted software called Think!Think! The primary objective of Think!Think! is to develop foundational math skills for elementary school students.

To examine the effect of Think!Think!, we ran a clustered randomized controlled trial (RCT) involving 1,656 students from grade 1 (G1) to G4 at five public elementary schools near Phnom Penh from May to August 2018. Because each school has two classes in each grade, students were randomly assigned during the 3-month intervention to either one of the 20 treatment classes that used Think!Think! or one of the 20 control classes.

Our results show that the average treatment effects on cognitive skills measured by several types of math achievement tests and intelligence quotient (IQ) tests are positive and statistically significant. The size of the effect is large, especially compared with previous studies conducted in developing countries: our study's preferred point estimates are 0.68–0.77 standard deviation for student achievements and 0.66 standard deviation for IQ scores, even after controlling for prior scores in the baseline survey, gender, grade, birth month, parental education,

and schools' time-invariant characteristics. Furthermore, the CAI-based software raises students' subjective expectations of attending college in the future. However, there is no significant effect on noncognitive skills, namely motivation and self-esteem.

Our contribution to the literature can be mainly summarized as follows: (i) While prior literature has focused more on test score gains, our paper examines the effect of CAI on a wide variety of outcome variables, including cognitive skills measured by test scores and IQ scores, noncognitive skills measured by motivation and self-esteem, and other habits such as hours spent studying at home. (ii) While the demand for new technologies is growing in education, especially in developing countries, more rigorous research is required to establish their external validity. To our knowledge, ours is the first study that has been rigorously designed and implemented in Cambodia. (iii) Unlike prior literature which provided after-school CAI as part of remedial education, students in our study were allowed to access CAI only during class. We are thus able to identify whether CAI caused an improvement in students' cognitive abilities because of increased learning productivity per hour, not because of increased hours available for instruction.

The remainder of this paper will proceed as follows. Section II provides a literature review. Section III explains the research design and data. Section IV presents empirical specifications and the main results on cognitive and noncognitive skills. Section V concludes and provides policy implications.

II. Literature Review

Previous studies have defined investment in computers by schools as either information communication technology or CAI. In recent years, CAI programs, which do not necessarily require an internet connection, have become more widely used in public schools. However, while several studies have shown that well-designed CAI programs appear to have strong and positive effects on math or science abilities of weaker students, especially in developing countries, other studies have found insignificant effects on reading and language test scores. For example, Rouse and Krueger (2004) ran a large-scale RCT using the computer software program Fast For Word for G3 to G6 students in an urban district in northwestern United States. Their results showed that the effect of this program on language and reading skills is small and statistically significant. Banerjee et al. (2007) examined the effect of a CAI program for G4 students in urban India. The students who were randomly assigned to treatment schools increased their math achievements by 0.47 standard deviation, mainly because of improvement among poorer performing children. Surprisingly, this positive effect remained even after the programs were terminated, although the size of the effect decreased to about 0.10 standard deviation.

In economics, investments in computers, the internet, software, and other technologies have been analyzed typically in the context of an education production function. Bulman and Fairlie (2016) pointed out that the binding constraint in the model is often the amount of time available for instruction, which is regarded as one of the educational inputs. In other words, this trade-off between time spent using a computer in class and time spent on traditional instruction makes it more difficult to determine whether schools should use CAI programs or more traditional instruction. However, many studies, including Rouse and Krueger (2004) and Banerjee et al. (2007), have estimated the effect of supplemental education or remedial education with CAI programs outside of class.

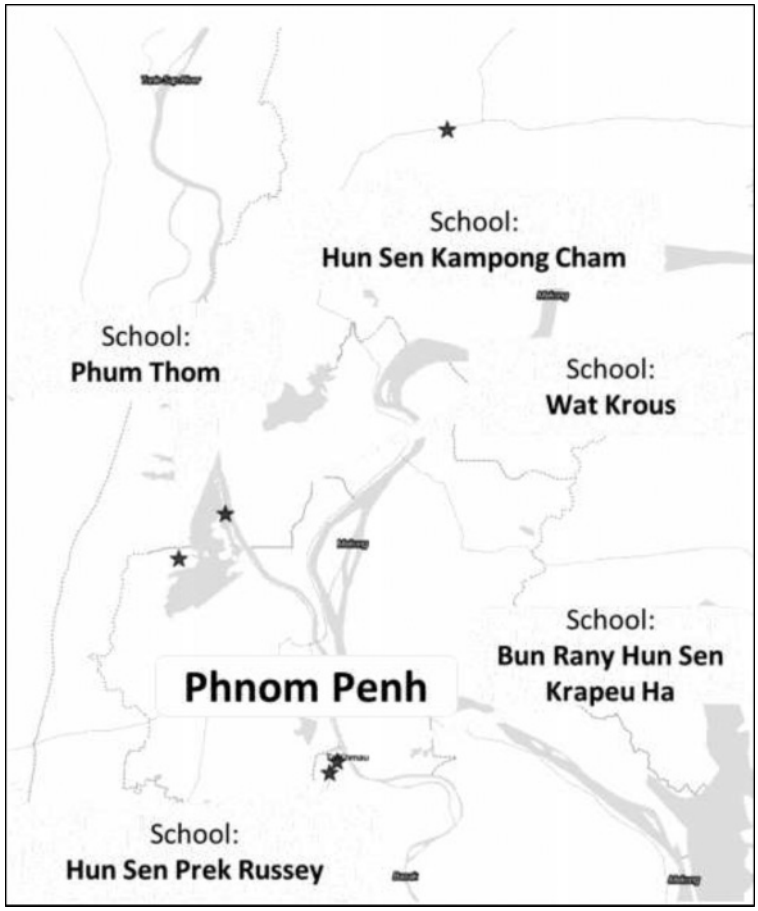
To deal with these issues, Barrow, Markman, and Rouse (2009) developed a trial in which middle school students in randomly assigned treatment classes were taught using CAI, while students in the control classes were taught traditionally in class. This enabled a comparison of the effects of the newly developed CAI program and more traditional instruction under limited school resources and time constraints. The 2-year experiment found that the treatment students improved their math ability by at least 0.17 standard deviation more than their counterparts. Carrillo, Onofa, and Ponce (2011) conducted a similar experiment in Ecuador for elementary school students. Using CAI in class, instead of traditional instruction, helped to improve math performance, but not language acquisition. However, a recent study on middle schools in urban India showed that using CAI in class has a greater impact on both math and language abilities (Muralidharan, Singh, and Ganimian 2019). The authors' instrumental variable estimates find that treatment students performed 0.37 standard deviation higher in math and 0.23 standard deviation higher in Hindi during the 5-month intervention. They also found that the achievement gains were greater for academically weaker students. Our empirical analysis follows that of Muralidharan, Singh, and Ganimian (2019) and tests whether CAI programs are effective for younger children in relatively disadvantaged areas of a developing country.

III. Methodology and Data

A. Background

Our study targets five public elementary schools located within a radius of approximately 10 kilometers around Phnom Penh. Because these schools did not receive any aid or assistance from other development agencies during the period of our intervention, we can rule out any confounding factors from other external interventions. The majority of households around the schools engage in farming and fishing to generate income. Only a small proportion of parents have tertiary education. The locations of these five schools are illustrated in Figure 1.

Figure 1. **Locations of Target Schools**



Note: This map was not produced by the cartography unit of the Asian Development Bank. The boundaries, colors, denominations, and any other information shown on this map do not imply, on the part of the Asian Development Bank, any judgment on the legal status of any territory, or any endorsement or acceptance of such boundaries, colors, denominations, or information.
Source: Stamen Maps. http://maps.stamen.com/.

B. Baseline and Follow-Up Surveys

Prior to the intervention, we conducted baseline surveys in class from 21–25 May 2018 with the full cooperation of teachers and staff. The baseline survey included two sets of 40-minute achievement tests for G3 and G4 students, 40-minute IQ tests for all students, and 20-minute surveys for all students and parents.

To measure students' cognitive skills, two sets of achievement tests were used: the National Assessment Test (NAT) administered by Cambodia's Ministry of Education, Youth and Sports for G3 students; and Trends in International

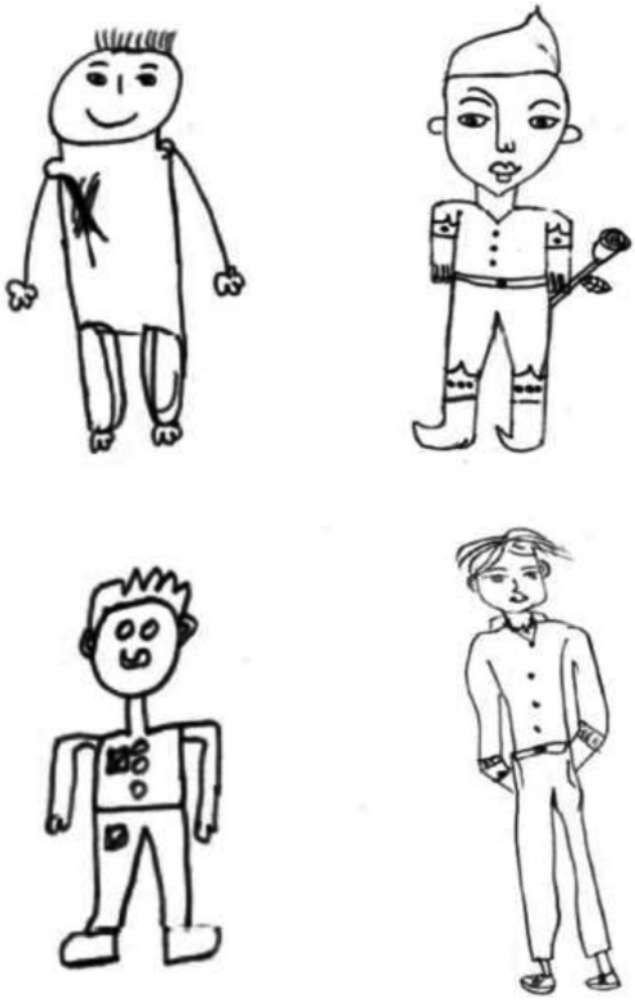
Mathematics and Science Study (TIMSS) administered by the International Association for the Evaluation of Educational Achievement (IEA) for G4 students. We selected exams that the students in our intervention had not previously taken. As there are no standardized tests to measure the cognitive abilities of younger students, we did not administer achievement tests for G1 and G2 students. Instead, we administered two sets of age-appropriate IQ tests in the baseline survey. One of the IQ tests—the "new Tanaka B-type intelligence test" (Tanaka, Okamoto, and Tanaka 2003)—has long been used in Japan and other countries in Asia as an age-appropriate measure of children's cognitive skills. The Tanaka B-type intelligence test was translated into the local language and also modified appropriately for the local environment (e.g., illustrations of local banknotes, food, people, etc.). The other intelligence test conducted during the baseline survey was the Goodenough Draw-a-Man (DAM) test (Goodenough 1926). In this test, students are asked to complete drawings of a whole person(s) on a piece of paper for 10–15 minutes. Several examples of children's drawings collected during our baseline survey appear in Figure 2. Although the validity of this test as a measure of intelligence has been criticized, the literature suggests that the DAM test is effective in screening for lower levels of intelligence in 5- to 12-year-old children (Scott 1981).

The survey of all G1 to G4 students asked them to provide demographic information, including gender, grade, birth month, hours of study at home, and subjective likelihood of attending college in the future. The survey also included a set of questionnaires to measure noncognitive skills, in particular the Rosenberg self-esteem scale (Rosenberg 1965) and an internal and external motivation scale (Sakurai and Takano 1985). The survey of parents asked about socioeconomic status, such as their educational backgrounds.

Following the 3-month intervention, a follow-up survey was conducted from August 16 to 25. We again administered the same sets of achievement tests, IQ tests, and questionnaires for students, focusing only on time-varying variables, such as willingness to attend college and time spent studying at home.

Out of 1,656 students who officially registered in our target schools, 77.2% of them participated both in the baseline and follow-up surveys, although 6.3% did the baseline survey only. The sample attrition may be a great threat to reduce the comparability of treatment and control. If our intervention is successful, the low-achieving students assigned to the treatment group may not drop out during the intervention, while their counterpart low-achieving students assigned to the control group may drop out of school altogether. In this case, the estimated impact of this intervention may be downward biased. We calculated the attrition rate for both treatment and control groups and checked whether the students who dropped out of the two groups had different characteristics. Fortunately, there is no evidence of differential attrition rates and different types of attrition in the treatment and control groups. However, we still do not know much about the 9.2% of students who

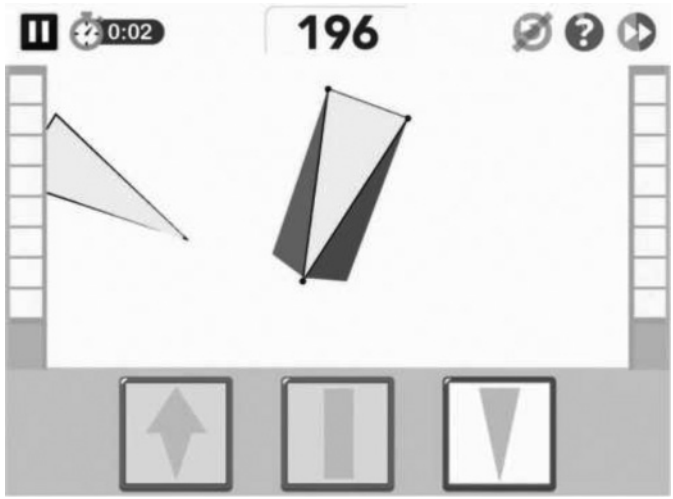
Figure 2. **Samples of Draw-a-Man Test**



Source: Figures by four anonymous students selected from the Draw-a-Man tests conducted during the baseline survey.

completed neither baseline nor follow-up surveys. According to the latest World Bank Indicators, the school dropout rate in Cambodia nationwide was 9.4% in 2017. Because our intervention was implemented in the last 3 months of the semester, some may have dropped out of school before or during the intervention. To deal with this problem, we created a dummy variable which we set to 0 if the baseline data is missing and then controlled for it in our analysis of covariance (ANCOVA) estimate as a robustness check (models 2 and 3 in Tables 2 and 3, and model 2 in Tables 4 and 5).

Figure 3. **Sample Problem**



Source: Wonder Lab (formerly Hanamaru Lab). https://wonderlabedu.com/.

C. Education Software: Think!Think!

The software called Think!Think! used in our intervention was originally developed by Hanamaru Lab, taking full advantage of its substantial experience in operating a large number of cramming schools for school-aged children in Japan. This software is especially designed to develop the foundational math skills of elementary school students (Figure 3). Why math? As a number of studies have suggested, math and science skills are highly related to economic growth across countries (e.g., Jamison, Jamison, and Hanushek 2007; Hanushek and Woessmann 2016). The benefits of mathematical proficiency not only drive economic growth but also raise individual earnings. For example, Joensen and Nielsen (2009) exploited an institutional reduction in the costs of acquiring advanced high school math in Denmark and provide evidence that the choice of a more math-intensive high school specialization has a causal effect on future labor market earnings. More specifically, Think!Think! incorporates adaptive learning using an original algorithm and provides math problems, materials, and instructions to reflect the proficiency level of each individual student.

Think!Think! was modified for elementary school students in Cambodia to meet local curriculum standards and was translated into the local language, Khmer. Students who were assigned to treatment classes were provided with free access to a tablet or laptop to use Think!Think! in class. CAI often requires additional teaching staff in class. In our intervention, we provided three additional staff with no teaching experience to advise students on technical matters and time management.

We carefully compared Think!Think! with a CAL program called Mindspark used in a study by Muralidharan, Singh, and Ganimian (2019) and found that it had many features that were very similar to Think!Think! According to the authors, the advantages of using Mindspark are (i) its high-quality instructional materials; (ii) its adaptive contents which allow them to implement "Teaching at a Right Level" for each individual student; (iii) that it alleviates a student-specific conceptual bottleneck; and (iv) its interactive user interface, all of which also characterize the attractive features of Think!Think! (Muralidharan, Singh, and Ganimian 2019, 1431–32). One slight difference is that Mindspark provides Hindi (language) programs as well as math for middle school students (G6 to G9), while Think!Think! specializes in math for younger primary school students (G1 to G4).

Because of these similarities, our results are, in fact, very consistent with Muralidharan, Singh, and Ganimian (2019). However, one of the most significant differences between our study and Muralidharan, Singh, and Ganimian (2019) is in the implementation. The authors' intervention was a "blended learning" program, meaning "a combination of the Mindspark CAL program, group-based instruction and extra instructional time" (Muralidharan, Singh, and Ganimian 2019, 1429). Their results, therefore, could not disentangle the pure effect of CAL from additional inputs and investigate whether the technology could have a positive effect on test scores in the absence of a constraint. Assuming that the amount of time available for instruction is fixed at a school, whether schools choose the optimal level of technology relative to traditional instruction in class may be a more relevant policy question for governments in developing countries.

D. Clustered Randomized Controlled Trial

If we were to allow students to access the CAI based on their own preferences, the software would most likely be used by higher-achieving students. Students who have sought to access a higher quality of education, including the exposure to new technology, are much more enthusiastic to study, on average, than those who never did. Random assignment of access to the CAI-based software avoids this selection bias.

Students in the treatment classes used Think!Think! for approximately 30 minutes each day. Peer effects are a potential threat to the internal validity of this experiment, and interactions between students may violate the stable unit treatment value assumption. To avoid this situation, besides the fact that clustered RCT is more common in education as noted in the literature, we randomized classrooms rather than individual students within them.[1]

[1]However, as pointed out by Imbens and Wooldridge (2009), it is technically difficult to separate the direct effect of the intervention on an individual from the indirect effect of peers on that individual.

Because each school has two classes in each grade, we used a stratified randomization: we picked one treatment in each grade at each school. This created 20 treatment classes (with 840 students) and 20 control classes (with 816 students) across the five schools.[2] However, there is still the concern that students in the treatment classes would talk to their friends in the control classes at the same school about what they had learned. To reduce the risk of such spillovers, we did not allow the treatment students to access Think!Think! outside of class. Furthermore, they were not allowed to take their tablet or laptop home. However, our class-level clustered randomization may not be enough to contain the spillovers between treatment and control groups. The unbiased estimate may be larger if there exists a positive spillover within treated peers and a secondary effect on those who are not treated in the same schools. Despite the relatively short period of intervention of 3 months, the students were enthusiastic about using Think!Think! The drawbacks of our study may be the presence of evaluation-driven behavioral changes in the treatment group called the Hawthorn effect and/or in the control group called the John-Henry effect. Because the Hawthorn effect artificially improves student's outcomes in the treatment group, the impact of CAI compared to its true impact may be overestimated, although we do not find any significant change in motivation within the treatment group. On the other hand, the John-Henry effect boosts outcomes among students in the control group, which may underestimate the impact of CAI.

IV. Econometric Specification and Results

A. Econometric Specification

To identify the causal effect of using Think!Think!, we conduct ANCOVA using the following model and identify the effect of using CAI. Our equation of interest is

$$Y_{i,j,t} = \alpha + \beta T_{i,j,t} + \gamma Y_{i,j,t-1} + \delta MissingBaseline_{i,j,t} + X_{i,j,t^{\sigma}} + \in_{i,j,t} \quad (1)$$

where $Y_{i,j,t}$ is the outcome variable of student i in school j at time t. $T_{i,j,t}$ is access to CAI and the key independent variable of interest. $MissingBaseline_{i,j,t}$ is a dummy variable to indicate whether student i participated in the baseline survey or not. $X_{i,j,t}$ is a vector of control variables, while $\in_{i,j,t}$ is the idiosyncratic error term. $X_{i,j,t}$ includes the basic demographic controls, such as gender, grade, birth month, parental education, and school-grade time-invariant fixed effects. According to

[2] While there can be unobserved correlations between the outcomes of students in the same classroom, clustered standard errors can be used to correct for such correlations. However, we cannot calculate clustered standard errors because there are only 40 classrooms in our experiments and the calculation of this type of standard errors requires at least 42 clusters, as suggested by Angrist and Pischke (2008).

McKenzie (2012), ANCOVA is preferred for experimental designs, rather than the difference-in-difference approach, when the autocorrelation in outcome variables between the baseline and the follow-up survey is low. Because our data are only weakly autocorrelated, we apply ANCOVA for our estimation.

The crucial identifying assumption in this empirical model is that the relationship between exposure to the CAI-based software and students' unobserved ability is orthogonal to the error term, conditional on the controls. Under this assumption, the estimate of β in equation (1) can be interpreted as the causal impact of the CAI-based software on student outcomes.

B. Variable Definitions

Table 1 presents a balance check for the baseline survey. There is no statistically significant difference in the results of the NAT between the G3 students assigned to treatment classes and those assigned to control classes, although the G4 students in the control classes performed slightly better on the TIMSS than those in the treatment classes, even after controlling for school-by-grade fixed effects, following Bruhn and McKenzie (2009).

Another outcome variable is IQ test scores: the results of the Tanaka B-type IQ test and the DAM test are converted to a mental age, and the IQ scores are then calculated as mental age divided by chronological age multiplied by 100. According to the descriptive statistics, the mean of the Tanaka B-type IQ test score is 78.612 with a standard deviation of 13.451, and the mean of the DAM type IQ score is 0.692 with a standard deviation of 0.207. There is no statistically significant difference between the Tanaka B-type IQ test score and the DAM score.

The next set of outcome variables, measures of noncognitive skills, are coded as the mean of a set of questionnaires specific to self-esteem and motivation. The self-esteem measure is slightly higher for the treatment students, while the motivation measure is similar across the two groups of students. All cognitive and noncognitive outcome measures are normalized to a mean of 0 and a standard deviation of 1 when we run the regression analysis.

Willingness to attend college is measured on a 3-point scale (from 1 = not likely to 3 = very likely) based on students' subjective expectations. Hours spent studying at home is measured on a 6-point scale (from 1 = not at all to 6 = more than 4 hours). We set the minimum of this variable to 0 and the maximum to 4 hours, and then we took the median value for categories between 2 (less than 30 minutes) and 5 (2–3 hours) on the 6-point scale. The key independent variable of interest denoted by $T_{i,j,t}$ is a dummy variable coded as 1 if students are assigned to a treatment class and 0 otherwise.

The demographic variables denoted by $X_{i,j,t}$, such as gender, age, and parental educational backgrounds, are very similar between the treatment and control students. The variable on parental education represents the highest level of

Table 1. **Descriptive Statistics and Balance Test**

	All	Control (A)	Treatment (B)	Difference (B)-(A)
Achievement test (NAT, G3)	0.538	0.522	0.554	0.031
	(0.207, 356)	(0.198, 177)	(0.214, 179)	(0.039)
Achievement test (TIMSS, G4)	0.292	0.330	0.252	−0.067*
	(0.203, 347)	(0.187, 174)	(0.211, 173)	(0.035)
IQ test (Tanaka-B)	78.612	78.432	78.795	0.401
	(13.451, 1,385)	(13.131, 700)	(13.777, 685)	(1.647)
IQ test (Draw-a-Man)	0.692	0.678	0.705	0.026
	(0.207, 1,217)	(0.206, 594)	(0.207, 623)	(0.033)
Self-esteem	2.762	2.726	2.794	0.039
	(0.549, 1,150)	(0.596, 535)	(0.502, 615)	(0.043)
Motivation	0.656	0.652	0.660	0.01
	(0.142, 996)	(0.150, 471)	(0.133, 525)	(0.013)
Willingness to go to college	2.410	2.342	2.467	0.108
	(0.771, 1,051)	(0.809, 482)	(0.734, 569)	(0.109)
Minutes of studying at home per week	168.667	170.142	167.481	−3.111
	(117.005, 949)	(108.975, 423)	(123.173, 526)	(13.628)
Gender (male = 1, female = 0)	0.525	0.530	0.519	−0.01
	(0.500, 1,643)	(0.499, 813)	(0.500, 830)	(0.016)
Age	8.485	8.501	8.470	−0.034
	(1.553, 1,620)	(1.573, 803)	(1.535, 817)	(0.048)
Highest parental education				
College or graduate school	0.023	0.016	0.029	0.009*
	(0.149, 1,236)	(0.127, 610)	(0.167, 626)	(0.004)
High school	0.457	0.474	0.441	−0.028
	(0.498, 1,236)	(0.500, 610)	(0.497, 626)	(0.031)
Junior high school	0.299	0.292	0.305	0.011
	(0.458, 1,236)	(0.455, 610)	(0.461, 626)	(0.019)
Elementary school	0.220	0.215	0.225	0.009
	(0.414, 1,236)	(0.411, 610)	(0.418, 626)	(0.022)
No education (ref)	0.002	0.003	0.000	−0.004
	(0.040, 1,236)	(0.057, 610)	(0.000, 626)	(0.003)
Birth month				
Jan–Mar	0.234	0.223	0.245	0.018
	(0.423, 1,620)	(0.416, 803)	(0.430, 817)	(0.021)
Apr–Jun	0.246	0.263	0.230	−0.034
	(0.431, 1,620)	(0.440, 803)	(0.421, 817)	(0.022)
Jul–Sep	0.249	0.255	0.244	−0.016
	(0.433, 1,620)	(0.436, 803)	(0.430, 817)	(0.019)
Oct–Dec	0.270	0.259	0.282	0.021
	(0.444, 1,620)	(0.438, 803)	(0.450, 817)	(0.021)

CAI = computer-aided instruction, G3 = grade 3, G4 = grade 4, IQ = intelligence quotient, NAT = National Assessment Test, TIMSS = Trends in International Mathematics and Science Study.

Notes: Treatment and control refer to whether students are randomly assigned into classes with CAI. Variables used in this table are from the baseline survey in May 2018. The data are combined from three pieces of survey conducted: (i) student survey, (ii) parent survey, and (iii) skill assessment. The numbers reported in each cell represent means along with the standard deviation and the number of observations in parentheses (in this order). The column "Difference" shows the estimates drawn from regressing outcomes on a treatment dummy coded 1 if students are randomly assigned into classes with CAI and school-by-grade fixed effects. ***, **, and * represent 0.1%, 1%, and 5% significance levels, respectively.

Source: Authors' estimates.

education of either one of the parents. Note that this information is retrieved from the parental survey conducted at the same time as the student survey. However, unlike the 100% response rate of the student survey administered during class, the response rate of the parental survey was approximately 85%.

Although the observable characteristics are similar between the two groups, several outcome variables, namely the achievement score for G4 students, DAM type IQ scores, and self-esteem scale, are not comparable in the baseline survey.

Because heterogeneity across groups can occur by chance even when randomization is implemented correctly and the chance of achieving homogeneity when we randomize at the group level increases with sample size, we are not concerned by heterogeneity in four of the 15 variables. However, although schools randomize the change in class composition annually, heterogeneity between the treatment and control groups may still exist because of dropouts or absences on the day of the baseline survey. We thus control for this using the demographic variables we use for the heterogeneity check to enable a "pure" comparison.

The average treatment effect may depend on the interests of particular subgroups of students. For example, if boys are more familiar with computer-related equipment, the effect may be stronger for boys than girls. This kind of heterogeneous effect is important for policy makers in designing policy to reflect the needs of particular subgroups. We will discuss this point in the next section.

C. Results

1. Effect on Cognitive Skills

We start by estimating the effect of CAI on student achievement. The ordinary least squares estimates are reported in Table 2 along with heteroskedasticity-robust standard errors. Our primary focus is the estimated effect of access to Think!Think! on the NAT for G3 students and on the TIMSS for G4 students in the first row of the table.

Model 1 provides unconditional ANCOVA estimates. Model 2 controls for prior achievement scores in the baseline survey and the missing baseline dummy. Model 3 controls for basic demographic controls, such as gender, grade, birth month, parental education, and school-grade time-invariant fixed effects, in addition to prior test scores and the missing baseline dummy.

The results clearly show that the estimated coefficients on the standardized test scores are positive and statistically significant at the 0.1% level (Table 2, NAT). The estimated coefficients for the sample of G3 students indicate that exogenous exposure to the CAI raises average test scores by about 0.77 standard deviation in model 3.

Table 2. **Effect of Treatment: Cognitive Skills**

	NAT			TIMSS		
Dependent Variable	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Treatment	0.814***	0.723***	0.767***	0.522***	0.630***	0.681***
	(0.291)	(0.204)	(0.223)	(0.135)	(0.091)	(0.104)
Baseline score		√	√		√	√
Control			√			√
Observations	369	369	298	350	350	303
Adjusted R^2	0.131	0.619	0.695	0.051	0.096	0.213

NAT = National Assessment Test, TIMSS = Trends in International Mathematics and Science Study.
Notes: The coefficients for the treatment group are reported above. The unit of observation is student. Columns labeled models 1–3 show ordinary least squares estimates. Model 2 controls for prior score and missing baseline dummy. Model 3 controls for prior score, gender, grade, birth months, parental education, missing baseline dummy, and school-grade fixed effects. Standard errors are in parentheses and clustered by school. ***, **, and * represent 0.1%, 1%, and 5% significance levels, respectively.
Source: Authors' estimates.

Adding demographic controls and school-by-grade fixed effects to model 3 neither changes the magnitude of the coefficients across specifications nor improves the precision of our estimates in explaining the variation in test scores. Once we include the interaction term and test for heterogeneous effects for gender, grade, and parental education, we obtain small point estimates on nearly all the interaction terms, and the differences between these coefficients do not support the hypothesis of significant heterogeneous effects on test scores. Furthermore, the achievement gains are homogeneous for academically weaker students. These results are available upon request.

The results are consistent with our expectations for the G4 sample (Table 2, TIMSS). Access to the CAI improves standardized test scores by 0.68 standard deviation per 3-month exposure in model 3. Adding controls increases the point estimates and decreases the standard errors of these estimates. At the same time, we do not find any significant heterogeneous effects of gender, grade, parental education, or initial achievement on test scores.

In Table 3, the estimated coefficient on the Tanaka B-type IQ score is positive and statistically significant at the 0.1% level. Table 3 shows that the effect on the IQ score from model 3 is 0.66 standard deviation. The estimated coefficient is unchanged after controlling for demographic characteristics in model 3. However, the coefficients of the DAM score are not statistically significant, regardless of the model specification. Overall, our results indicate that the magnitude in cognitive skills appears to be very large, compared with evidence from previous literature where the intervention lasted for at least a year.

Because Muralidharan, Singh, and Ganimian (2019) applied very similar CAL software to relatively poor students in Delhi, India, it is worth comparing their results with ours. The comparable intent-to-treat estimates in Muralidharan, Singh, and Ganimian (2019) indicate that lottery-winner-treated students scored 0.23

Table 3. **Effect of Treatment: Intelligence Quotient (IQ)**

	IQ			Draw-a-Man		
Dependent Variable	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Treatment	0.705***	0.692***	0.664***	0.071	0.022	−0.003
	(0.117)	(0.111)	(0.111)	(0.080)	(0.097)	(0.105)
Baseline score		√	√		√	√
Control			√			√
Observations	1,404	1,404	1,146	1,390	1,390	1,133
Adjusted R^2	0.076	0.4	0.51	0.001	0.182	0.287

Notes: The coefficients for the treatment group are reported above. The unit of observation is student. Columns labeled models 1–3 show ordinary least squares estimates. Model 2 controls for prior score and missing baseline dummy. Model 3 controls for prior score, gender, grade, birth months, parental education, missing baseline dummy, and school-grade fixed effects. Standard errors are in parentheses and clustered by school. ***, **, and * represent 0.1%, 1%, and 5% significance levels, respectively.
Source: Authors' estimates.

standard deviation higher than control students after 4.5 months, while our results in model 3 show an improvement of 0.77, 0.68, and 0.66 standard deviation for G3 students who took the NAT, G4 students who took the TIMSS, and G1–G4 students who took the Tanaka-B IQ test, respectively. Muralidharan, Singh, and Ganimian (2019) recruited the sample students from a cramming school called Mindspark center in Delhi, and parents were told that their children would be chosen by lottery to receive a tuition waiver (₹200 per month, equivalent to $3). Their participants were self-selected (and perhaps highly motivated) and the administrative data suggested they performed better than nonparticipants. Muralidharan, Singh, and Ganimian (2019) found considerable heterogeneity in student progress by initial learning level and that test score gains were much larger for initially low-achieving students. The true estimates drawn from the representative sample containing more low-performing students in our study may be much larger than the estimates reported in their paper. On the other hand, because we covered all students in public schools, the participants were not self-selected into the intervention.

Using kernel density estimation, we obtain the probability density function for both achievement test scores and IQ scores to compare the score distributions after the 3-month intervention (Figures A1–A3). Although the difference in the DAM scores for the entire sample and even the interaction term with grades are not statistically significant, the skills of younger students seem to improve.

2. Effect on Noncognitive Skills and Inputs for Study

We then repeated the above approach using a set of noncognitive skills as outcomes. Unlike the results for cognitive skills, we do not find any significant effect for noncognitive skills, measured by motivation and self-esteem (Table 4). However, it is clear that the estimated coefficient on willingness to attend college is positive and statistically significant at the 5% level (Table 5), indicating that

Table 4. **Effect of Treatment: Noncognitive Skills**

	Motivation		Self-Esteem	
Dependent Variable	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Treatment	−0.023	−0.031	0.023	0.014
	(0.070)	(0.069)	(0.052)	(0.059)
Baseline score	√	√	√	√
Control		√		√
Observations	1402	1125	1396	1121
Adjusted R^2	0.274	0.377	0.025	0.138

Notes: The coefficients for the treatment group are reported above. The unit of observation is student. Columns labeled models 1 and 2 show ordinary least squares estimates. Model 1 controls for prior score and missing baseline dummy. Model 2 controls for prior score, gender, grade, birth months, parental education, missing baseline dummy, and school-grade fixed effects. Standard errors are in parentheses and clustered by school. ***, **, and * represent 0.1%, 1%, and 5% significance levels, respectively.
Source: Authors' estimates.

Table 5. **Effect of Treatment: Study Input**

	Study Time (minutes)		Willingness to Go to College	
Dependent Variable	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Treatment	−0.032	−0.099	0.136*	0.139*
	(0.097)	(0.101)	(0.073)	(0.083)
Baseline score	√	√	√	√
Control		√		√
Observations	1,299	1,057	1,367	1,094
Adjusted R^2	0.05	0.09	0.033	0.048

Notes: The coefficients for the treatment group are reported above. The unit of observation is student. Columns labeled models 1 and 2 show ordinary least squares estimates. Model 1 controls for prior score and missing baseline dummy. Model 2 controls for prior score, gender, grade, birth months, parental education, missing baseline dummy, and school-grade fixed effects. Standard errors are in parentheses and clustered by school. ***, **, and * represent 0.1%, 1%, and 5% significance levels, respectively.
Source: Authors' estimates.

students who used the CAI during class are more likely to believe they would undertake more advanced education in the future. The coefficient remains constant after controlling for demographic characteristics in model 2, which suggests that heterogeneous effects in terms of gender, grade, and parental education do not exist. Although the results do not indicate a positive effect of the CAI on noncognitive skills, the estimated probability density functions (Figures A4–A5) suggest a slight improvement in younger grades.

We also estimated the effect on time spent studying at home (Table 5), which is considered an important input of an education production function. As already mentioned above, students were not allowed to bring the tablet or personal computer to their own homes. It is thus convincing that we do not find any significant effect on studying longer at home. However, students in treatment classes sharply raised

their achievements, even though their hours of study did not change both at home and in the classroom. This indicates that CAI is successful in improving students' learning efficiency and productivity.

V. Conclusion

We examined the causal effect of CAI on children's cognitive and noncognitive skills. In collaboration with the Government of Cambodia, we ran a clustered RCT at five elementary schools around Phnom Penh over a period of 3 months. Students were randomly assigned to either one of 20 treatment classes that were allowed to use the CAI instead of regular math classes during the intervention or one of 20 control classes. Our empirical results show that the average treatment effect on cognitive skills measured by several types of math achievement tests and IQ tests is positive and statistically significant. The effect size is large, especially compared with those in previous studies for developing countries: the estimated coefficients are 0.68–0.77 standard deviation for student achievement and 0.66 standard deviation for IQ scores even after controlling for demographic factors. Furthermore, we found that the CAI can raise students' subjective expectation of attending college in the future. However, there is no significant effect on noncognitive skills, namely motivation and self-esteem.

Because we ran this clustered RCT for only 3 months, whether these effects remain in the longer term requires further investigation. Nevertheless, our results suggest that CAI has tremendous potential to improve students' math scores in both the short term and possibly the longer term.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235–64.

Barrera-Osorio, Felipe, and Leigh L. Linden. 2009. "The Use and Misuse of Computers in Education: Evidence from a Randomized Controlled Trial of a Language Arts Program." *Abdul Latif Jameel Poverty Action Lab (JPAL)*. https://www.povertyactionlab.org/evaluation/use-and-misuse-computers-education-evidence-randomized-controlled-trial-language-arts.

Barrow, Lisa, Lisa Markman, and Cecilia Elena Rouse. 2009. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." *American Economic Journal: Economic Policy* 1 (1): 52–74.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1 (4): 200–32.

Bulman, George, and Robert W. Fairlie. 2016. "Technology and Education: Computers, Software, and the Internet." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 5: 239–80. Amsterdam: Elsevier.

Carrillo, Paul E., Mercedes Onofa, and Juan Ponce. 2011. "Information Technology and Student Achievement: Evidence from a Randomized Experiment in Ecuador." IDB Working Paper No. 78.

Cristia, Julian, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Sever'in. 2017. "Technology and Child Development: Evidence from the One Laptop per Child Program." *American Economic Journal: Applied Economics* 9 (3): 295–320.

De Melo, Gioia, Alina Machado, and Alfonso Miranda. 2014. "The Impact of a One Laptop per Child Program on Learning: Evidence from Uruguay." IZA Discussion Paper No. 8489.

Glewwe, Paul, and Karthik Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 5: 653–743. Amsterdam: Elsevier.

Goodenough, Florence Laura. 1926. *Measurement of Intelligence by Drawings.* Yonkers-on-Hudson, New York: World Book Company.

Hanushek, Eric A., and Ludger Woessmann. 2016. "Knowledge Capital, Growth, and the East Asian Miracle." *Science* 351 (6271): 344–45.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Jamison, Eliot A., Dean T. Jamison, and Eric A. Hanushek. 2007. "The Effects of Education Quality on Income Growth and Mortality Decline." *Economics of Education Review* 26 (6): 771–88.

Joensen, Juanna Schrøter, and Helena Skyt Nielsen. 2009. "Is There a Causal Effect of High School Math on Labor Market Outcomes?" *Journal of Human Resources* 44 (1): 171–98.

Linden, Leigh L. 2008. "Complement or Substitute? The Effect of Technology on Student Achievement in India." InfoDev Working Paper No. 17. World Bank Group.

McKenzie, David. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–21.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109 (4): 1426–60.

Rosenberg, Morris. 1965. *Society and the Adolescent Self-Image*. New Jersey: Princeton University Press.

Rouse, Cecilia Elena, and Alan B. Krueger. 2004. "Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically Based' Reading Program." *Economics of Education Review* 23 (4): 323–38.

Sakurai, Shigeo, and Seijun Takano. 1985. "A New Self-Report Scale of Intrinsic versus Extrinsic Motivation toward Learning in Children." *Tsukuba Psychological Research* 7: 43–54.

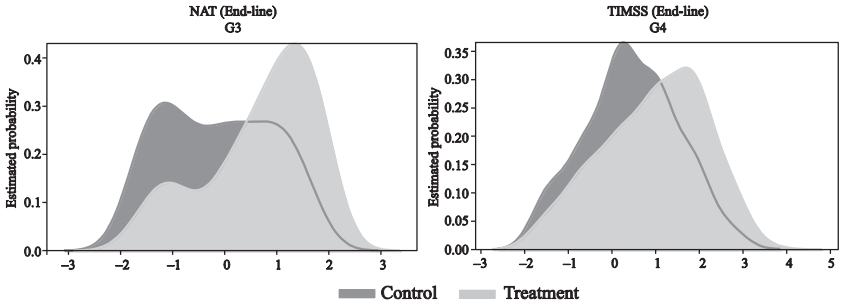
Scott, Linda H. 1981. "Measuring Intelligence with the Goodenough-Harris Drawing Test." *Psychological Bulletin* 89 (3): 483–505.

Tanaka, Kanichi, Kenroku Okamoto, and Hidehiko Tanaka. 2003. *The New Tanaka B Intelligence Scale*. Tokyo: Kaneko shobo.

World Bank. 2017. *World Development Report 2018*: *Learning to Realize Education's Promise*. Washington, DC.

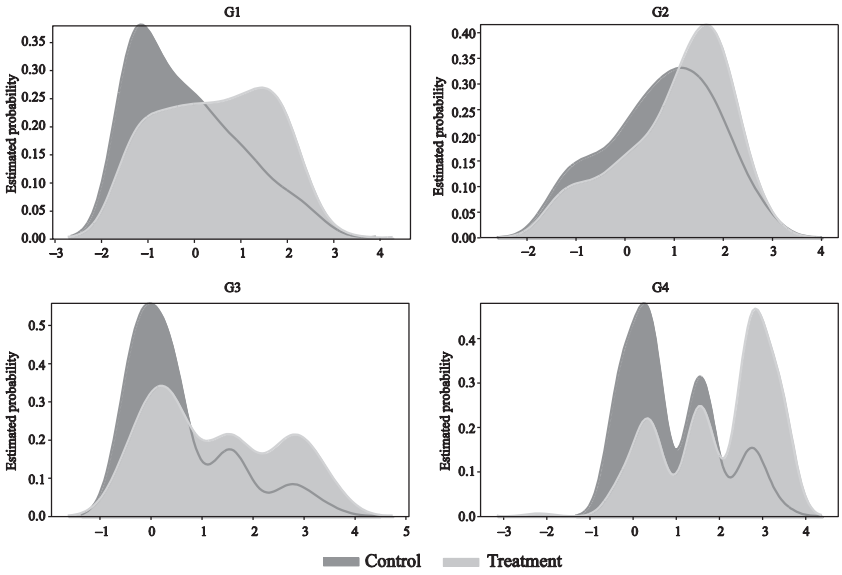
Appendix. Effect of Treatment: Estimated Probability Density Functions

Figure A1. **NAT and TIMSS Scores**



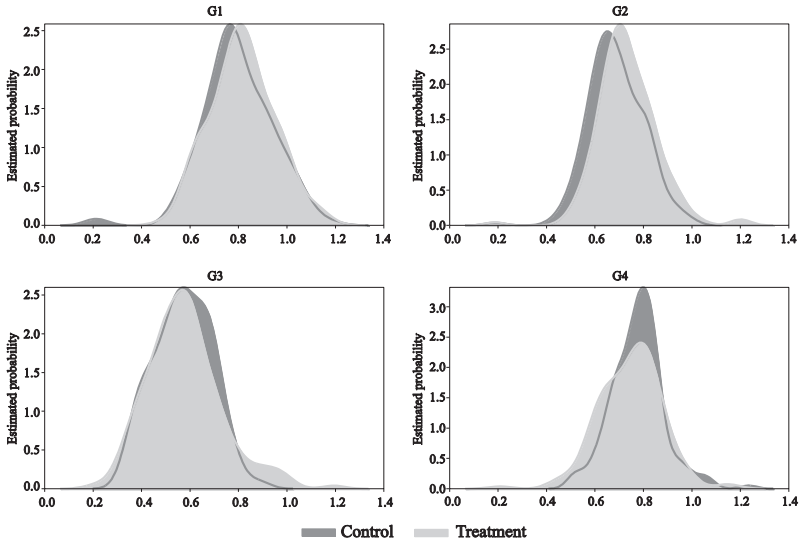
G3 = grade 3, G4 = grade 4, NAT = National Assessment Test, TIMSS = Trends in International Mathematics and Science Study.
Notes: This graph shows the estimated probability density functions for the National Assessment Test (NAT) and Trends in International Mathematics and Science Study (TIMSS) test given at the follow-up surveys. The light gray function represents treatment groups and the dark gray function represents control groups.
Source: Authors' calculation.

Figure A2. **Intelligence Quotient (IQ) Scores (End-line)**



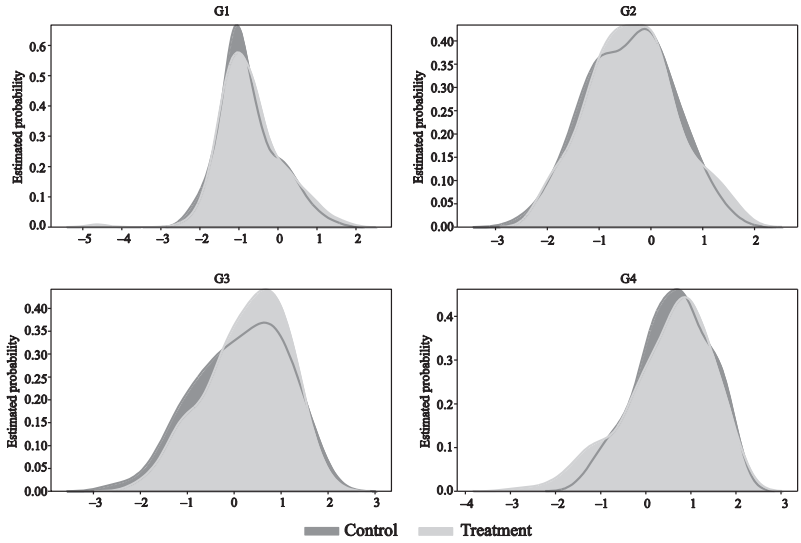
G1 = grade 1, G2 = grade 2, G3 = grade 3, G4 = grade 4.
Notes: This graph shows the estimated probability density functions for the IQ tests given at the follow-up surveys. The light gray function represents treatment groups and the dark gray function represents control groups.
Source: Authors' calculation.

Figure A3. **Draw-a-Man Test Scores (End-line)**



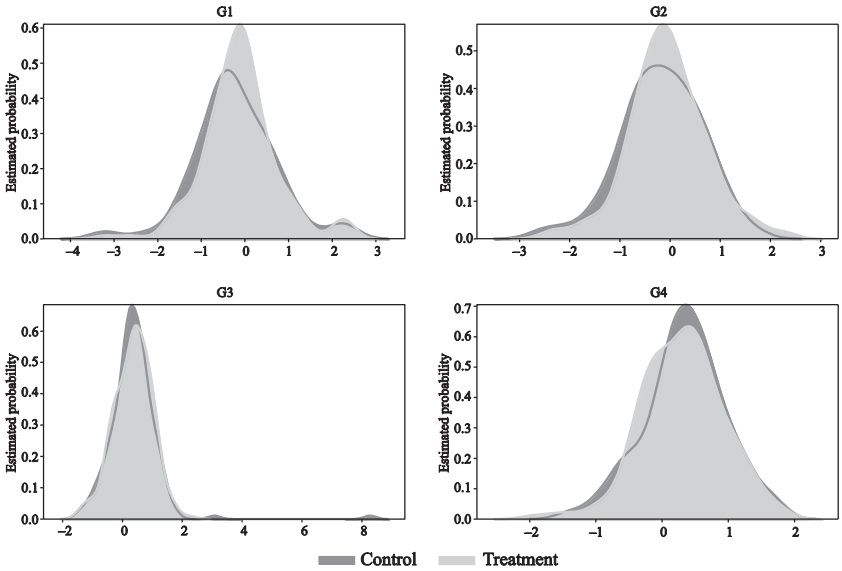
G1 = grade 1, G2 = grade 2, G3 = grade 3, G4 = grade 4.
Notes: This graph shows the estimated probability density functions for the Draw-a-Man test given at the follow-up surveys. The light gray function represents treatment groups and the dark gray function represents control groups.
Source: Authors' calculation.

Figure A4. **Motivation (End-line)**



G1 = grade 1, G2 = grade 2, G3 = grade 3, G4 = grade 4.
Notes: This graph shows the estimated probability density functions for motivation measured at the follow-up surveys. The light gray function represents treatment groups, and the dark gray function represents control groups.
Source: Authors' calculation.

Figure A5. **Self-Esteem (End-line)**



G1 = grade 1, G2 = grade 2, G3 = grade 3, G4 = grade 4.
Notes: This graph shows the estimated probability density functions for self-esteem measured at the follow-up surveys. The light gray function represents treatment groups and the dark gray function represents control groups.
Source: Authors' calculation.